N73 22334

LARS Information Note 032973

CASE FILE COPY

# Urban Land Use Mapping by Machine Processing of ERTS-1 Multispectral Data: A San Francisco Bay Area Example

by
R. Ellefsen, P.H. Swain and
J.R. Wray

The Laboratory for Applications of Remote Sensing

Purdue University, West Lafayette, Indiana

1973

# Urban Land-Use Mapping by Machine Processing of ERTS-1 Multispectral Data: A San Francisco Bay Area Example*

Richard Ellefsen
Geography Department
California State
University, San Jose

Philip H. Swain
Laboratory for
Applications of
Remote Sensing
Purdue University

James R. Wray
U.S. Geological Survey
Department of the
Interior

## I. Introduction

Discussed here are the results of attempts to create computer-produced urban land-use maps using multispectral scanner data from a satellite. The study is an outgrowth of research questions posed by individuals connected with LARS/Purdue, NASA, the Earth Resources Observation Program of the Department of the Interior, and the Geographic Applications Program of the U.S. Geological Survey.

Specific study objectives have been: (1) by LARS to test the applicability of the LARSYS pattern recognition software to land-use studies in an area where contemporaneous ground truth was available; and (2) by the Census Cities Project of the Geographic Applications Program to attempt to utilize ERTS-1 data as a support or possible replacement for land-use mapping achieved through conventional air-photo interpretation. Further Geographic Applications Program goals are to maximally utilize the ERTS-1 data to (1) produce print-out maps of large scale (approximately 1:24,000); (2) create digitized land-use data which may be used in conjunction

*Extended summary submitted to the Conference on Machine Processing of Remotely Sensed Data, Purdue University, 1973. This research was supported in part by NASA Grant NGL 15-005-112.

with such reported ground-collected data as census reports and parcel ownership; and (3) monitor urban change on a regular (18-day cycle) basis.

Geographic Applications Program forecasts heavy use of the mapped data by land-use planners and others who require frequently updated digitized spatial land-use data. Users will require these data in map form and in digitized form, aggregated by census tracts and such other units as corporate cities, school, planning, fire, sewage, and conservation districts.

## II. Data Processing

This study employs some of the procedures used successfully by LARS in crop-identification and other remote sensing studies; in particular the use of training fields (the "supervised" approach) where the computer is "trained" to recognize the spectral characteristics of land uses, and a cluster analysis approach where the computer program defines its own classes in an "unsupervised" manner based on spectral differences inherent in the data.

In either instance, some ground-truth information is required. For the supervised approach, typical samples of each use under study are selected from the ground-truth material (in this instance, air-photo imagery) and supplied to the computer program. In the cluster approach, ground-truth is employed after-the-fact to identify what, in reality, the computer program has recognized as distinguishable patterns.

To this point in the study, the cluster analysis approach has been the most successful. Its use required several distinct steps. First, the computer was instructed to classify each resolution element (each representing a multispectral reading of an 80 meter square, approximately 1.1 acre, on the ground) as one of a given number (fifteen) of spectrally different classes. A print-out of the classes was compared against the imagery representing ground-truth to determine the nature of the land uses represented by each of the machine-derived spectral classes.

It was then possible to begin to develop a land-use classification scheme. Optimally, the scheme would maximize the inherent subtle distinctions of the machine's spectral classification while being compatible with the functional urban land-use classes familiar to the city planner and other users of urban data. Essentially, the problem was to translate into informational classes the spectral classes which resulted from cluster analysis of the data from the four ERTS-1 multispectral scanner channels.

The land-use classification system which has proven to date to be the best compromise between spectral and functional classes is comprised of the eight classes appearing in Figure 1.

A prime goal throughout the experiment has been to determine the applicability of a scheme for the classification of land-use based on remotely sensed data proposed by Anderson et al.[1] of

[1] James R. Anderson, Ernest E. Hardy, and John T. Roach, <u>A Land-Use Classification System for Use with Remote Sensor Data</u>. Geological Survey Circular 671, Washington, 1972.

Figure 1

Spectral Characteristics of Functional
Urban Land-Use Classes

San Jose, California Area

| Functional Land Use | Spectral Characteristics |
|---|---|
| Transportation | Highly reflective concrete surfaces and low reflective asphalt surfaces |
| Livelihood | Highly reflective bright roof surfaces and low reflective tar surfaces. Both have high roof to ground surface ratios within the resolution element |
| Pre-WW II Residential | Low Reflectivity of roofs, driveways, and sidewalks plus mature vegetation |
| Post-WW II Residential | Medium to high reflectivity of roofs, driveways and sidewalks with immature vegetation |
| Vacant Land (Unimproved Open Space) | Moderately high reflectivity of unirrigated grassland |
| Improved (irrigated) Open Space | Highly responsive in the infra-red bands |
| Water | Highly light absorbing |
| Miscellaneous uses which "thresholded" | Extraordinarily bright surfaces of varying land uses such as metal mobile home roof-tops, new concrete, and standing, ripe grain crops |

the U.S. Geological Survey. That system has levels of generalization I, II, III, and IV; the more discrete higher level uses are collapsible into the more general lower level classes. Comparison with the classification system developed in this experiment (see Figure 2) demonstrates, at least for the studied urban sample, that machine processing of satellite scanner data is capable of a much finer classification than simply Level I. Level II is achieved in the delimitation of residential use, some transportation, open space, grassland, water, and "thresholded" open space.* In addition, the very fact that we are dealing with spectral characteristics permits (even requires in order to maximize the available data) the division of residential into the two sub-classes of pre- and post-World War II, and to differentiate between improved and unimproved open space.

In short, the combination of the satellite-borne scanner and machine processing provides a different tool than either conventional air photo interpretation or surface and statistical unit mapping. Each produces a somewhat similar but yet different product and each has advantages and disadvantages. While these have been well-documented for air photography and ground methods, the characteristics of machine-processed satellite data as applied to urban mapping are not well-known and deserve presentation.

*"Thresholded" points represent resolution elements having spectral characteristics vastly different from the characteristics of any of the spectral classes used in the analysis. They usually occur because the ground cover type was not represented in the ground area used to "train" the classifier.

Figure 2

Comparison of a Land-Use Classification Derived from Automatic Machine Processing of ERTS-1 MSS Data with a U.S.G.S. Proposed Land-Use Classification System for Use with Remote Sensor Data

| Machine-processed ERTS-1 Data | U.S.G.S. Proposed System | |
|---|---|---|
| | Level I | Level II |
| 1. Pre-WW II Residential | 01. Urban and Built-up | 01. Residential |
| 2. Post-WW II Residential | " | " |
| 3. Livelihood | " | 02. Commercial and Services; 03. Industrial; 04. Extractive; 06. Institutional; 07. Strip and Clustered Settlement; 08. Mixed |
| 4. Transportation | " | 05. Transportation, Communications, and Utilities |
| 5. Improved (irrigated) Open Space | " | 09. Open and Other |
| 6. Vacant Land (unimproved open space) | 02. Agricultural | 01. Cropland and pasture<br>02. Orchards, Groves Bush Fruits, Vineyards, and Horticultural Areas |
| | 03. Rangeland | 01. Grass; 02. Savannas (Palmetto Prairies); 03. Chaparral; 04. Desert Shrub |
| 7. Water | 05. Water | 01. Streams and Waterways; 02. Lakes; 03. Reservoirs; 04. Bays and Estuaries; 05. Other |
| 8. Miscellaneous Uses which "Threshold" | 04. Forest land | 01. Deciduous; 02. Evergreen (Coniferous and Other); 03. Mixed |
| | 06. Nonforested Wetland | 01. Vegetated; 02. Bare |
| | 07. Barren Land | 01. Salt Flats; 02. Beaches; 03. Sand Other Than Beaches; 04. Exposed Rock<br>05. Other |
| | 08. Tundra | 01. Tundra |
| | 09. Permanent Snow, Icefields | 01. Permanent Snow, Icefields |

The advantages are:

1. High speed processing
2. Frequently obtained new data
3. Unbiased and uniformly repetitive classification
4. Production of print-out maps at a large map scale at relatively low cost (once operational)
5. The inherent digitizing of land-use data retrievable in virtually any form or combination of forms

The disadvantages are:

1. The inability of the system to discriminate with consistent success between functionally dissimilar but spectrally similar land uses
2. The impossibility of detecting parcel ownership
3. Generalization by resolution element: at 80-meter resolution the complexity of the urban landscape cannot be shown fully
4. Identifications dependent on vegetation vary seasonally
5. Uncontrollable incidence of cloud cover

Review of the above suggests that for many potential users, the satellite/machine-processing system has advantages which outweigh disadvantages and will be welcomed as a new, powerful tool in spatial analysis work. Other users may have to await refinements in the system which will surely come with subsequent developments in scanner and data processing capabilities.

## III. Test Area

The bulk of the tests to date have been made in the urbanized section at the southern end of San Francisco Bay in Santa Clara

County and comprising the cities of Palo Alto, Mountain View, Sunnyvale, Santa Clara, San Jose, and several lesser suburban municipalities. The entire area covers only 1.5 centimeters by 4.5 centimeters on the standard photographically-reproduced ERTS-1 frame (at a scale of 1:1,000,000). At the enlargement given by the computer print-out--at a scale of approximately 1:24,000--the mosaic of print-out sheets covers a width of some 88 inches (8 eleven inch wide sheets) by some 39 inches in the north-south axis.

The test area has many general characteristics in common with rapidly growing cities throughout the U.S. Development has been more horizontal than vertical with: large areas of new single-family residences built on converted agricultural land; several large clusters of new industrial complexes; shopping centers; and various institutions. The original core areas of the nuclei cities, from which growth spread in the past twenty years, remain as islands of old within the new but many of these have been significantly altered. Connecting ribbons of commercial development are visible even to the unaided eye on the ERTS-1 imagery. Small patches of unbuilt-upon agricultural remnants remain as enclaves while exclaves of the expanding city are found in the rural-urban interface area.

## IV. Results

A significant factor affecting results of this experiment is that the urban land-use mapping process requires the combining of several basically different ground cover classes into a

single land-use class.* Single-family residential use, for example, is composed of such spectrally diverse features as asphalt streets, concrete drives and patios, shake roofs, varying levels of maturity of landscaping, "corner" grocery stores, churches, and schools. In addition, regional environmental characteristics affecting landscaping plus local varieties of building and paving materials, serve to reduce identification reliability.

Despite these difficulties, tests of classification accuracy have shown a high degree of reliability within the fairly broad classes employed and for area which is predominantly urban. (Figures showing examples of results will be included in the complete paper.)

Successful classification is accounted for largely by the general tendency for a specific urban function to be conducted in a specialized urban environment which yields a unique spectral signature. Industry, for example, is nearly always found in buildings or clusters of buildings with flat roofs. These surfaces are spectrally quite distinct from a residential area with pitched roofs, landscaping, and a full network of accompanying streets. As a result, for an estimated ninety percent of the mapped urban area, land-use identifications are correct.

Where a less uniform pattern of symbols is seen on the printout map, ground truth examination reveals that such areas are

*In agricultural studies, by contrast, the usual goal is discrete crop identification. Except for the problems of seasonal change and variations of stand, agricultural land uses are reasonably homogeneous.

indeed quite diverse and present classification problems even to the land-use mapper on the ground. Cases of incorrect identification are anomalies within the type and frequently are simply cases of recent or on-going construction, areas which have not had sufficient time to weather into a more typical spectral signature.

Reliability does, of course, vary from one-land use to another. Improved (irrigated) open space is one-hundred percent correctly identified, thanks to its high infra-red response. Vacant land (unimproved open space) is almost as accurately identified due to lack of man-made structures upon it and the uniformly high reflectivity of mature grass or bare ground in the area in summertime.

The remaining land uses have a somewhat poorer performance record. Ready reasons appear for each case, raising hopes that with the problem well identified, either a statistical or an opto-electronic solution may be found. Generally, identification problems revolve around spectral diversity within a class: apparently, the greater the diversity, the greater the chance of incorrect classification. Figure 3, illustrating the relationships between the "spectral continuum" and land uses, indicates the degree of diversity for each land-use class.

Transportation, for instance, is composed of the spectral opposites of both highly reflective concrete surfaces--interstate highway interchanges, airport aprons--and light-absorbing asphalt highways, parking lots, and runways. The livelihood class is

Figure 3

Location of Components of Functional Land-Use Classes

Along the "Spectral Class Continuum"

| LT | LV | OU | OU | R2 | OP | R2 | LV | R2 | R3 | R3 | R3 | LT | R3 | LT |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |

Classes Produced by Unsupervised Cluster Analysis

| Land-Use | Classes |
|---|---|
| LT - Transportation | 1, 13, 15 |
| LV - Livelihood | 2, 8 |
| OU - Open Space, Unimproved | 3, 4 |
| OP - Open Space, Improved | 6 |
| R2 - Post WW II Residential | 5, 7, 9 |
| R3 - Pre WW II Residential | 10, 11, 12, 4 |

also composed of extremes of light and dark surfaces; large areas are under roof cover some of which is highly reflective, others are made of light-absorbing dull asphalt materials. Also occupying different segments of the classification continuum are older residential versus newer residential uses, the former generally "darker" than the latter.

A number of identification problems are common to all of the classes involving man-made cover of the land. Of greatest concern is that functional land-use is not consistently reflected in building shape or spectral characteristics. The user of urban land-use data has a real need to distinguish between such diverse functions as retail, education, wholesaling, and transportation. When all these functions are found in spectrally similar settings, discrimination using spectral information alone is impossible. Attempts to determine reliable signatures for commercial versus industrial have proven inconclusive: while commercial establishments, such as along arterials and in shopping centers, generally have asphalt roof surfaces, industries exhibit bright and dull surfaces in about equal quantity.

An added problem is the differential weathering of all types of man-made surface materials. Old and new paving and roofing materials are spectrally distinct enough for the computer to class them differently even though functional land use is the same throughout. While the ability of the computer to make this distinction suggests a potential to differentiate newer from older developments, there is no ready accounting for replaced or refurbished

surfaces within old area. There is the suggestion that an advanced state of weathering which is spectrally discernible is an indicator of maintenance neglect and therefore low quality. With the help of equally neglected landscaping, it is possible that poorer residential areas may be identified. Similar interpretation of commercial-industrial classes offers less promise.

The other major unsolved problem to date is the confusing spectral similarity between urban man-made land cover and certain natural cover. A frustrating example has been the confusion between some dark-roofed, clustered buildings and dark, brushy natural vegetation located in mountains at great distances from any built-up area.

## V. Conclusions and Forecast

Results of the experiment to date demonstrate that producing land-use maps of a large scale by machine processing of ERTS-1 scanner data is feasible. By keeping land-use classes fairly broad, a remarkable level of accuracy is attained despite the relatively coarse resolution and the inherent complexities of man-made land cover.

Of equal value, this experiment suggests the route for follow-on work. For instance, the reliability of land-use recognition and classification needs improvement. The task requires both the urban geographer and the computer scientist. The former would undertake a systematic study of relationships between spectral characteristics and functional land-use classes while the latter would explore ways

of directing machine capabilities to the problem of identifying urban land uses. A significant contribution here would be the development algorithms using context to solve certain classification problems. These techniques--somewhat analogous to identification by association procedures used in photo-interpretation--are required for the computer to make distinctions between such functionally different but spectrally similar land uses as factories and shopping centers.

A second major thrust must be made towards developing and fitting optimal classification schemes to the capabilities of scanner-produced data from Skylab, ERTS-B, and aircraft from varying altitudes. The larger map scale (and the smaller area of reach resolution element) of aircraft-borne scanners may increase the probability of use identification and lend insight to work with the ERTS-1 scale. The inclusion of a thermal band, as planned for ERTS-B, would add a useful variable.

Once each data source were optimized, further work would be required to tailor maps and accompanying data to various user levels. The entire range of users from local to national planning could be accommodated.

Adding further utility to the maps would be the superimposing of boundaries of varying areal ground data-reporting units. Possibilities range from geometric grids to man-delineated statistical areal units. The ability to aggregate land uses by such areas and to monitor change with great frequency holds enormous promise for

such valuable measurements as intercensal population estimates.

Essential too is the study of a temporal series of ERTS passes for the purpose of monitoring and detecting change. Many of the chores presently done by hand lend themselves to machine processing. Important products would be precise measurement of incremental growth of subdivision housing, commercial (livelihood), and transportation uses. Summary statements of the type of change from one use to another could be facilitated by machine processing. A further advantage would be the ability to detect the fairly small-area changes within the body of the old city at the large map scale of the computer print-out.

The solution of several other practical problems could be furthered by the use of a scanner/computer-analysis system. A relatively simple task would be the frequent up-dating of the boundary of a city's urbanized area, a service of great value in varying advanced planning needs. Commercial applications are also possible in such common jobs as selecting optimum locations for stores, banks, and service stations. The location and measurement of open space, a matter of key general concern, would also be easily handled by such a system. Another practical problem which could be dealt with is the required measurement of land-use, present and projected, as a basis for mass-transit planning. Also, as suggested earlier, careful work may yield a method for measuring housing quality.

In sum, the advantages of speed, relative low cost, and frequent synoptic monitoring could be of extreme value in helping to solve many land-use problems.